SK Telecom's 5G Architecture Design & Implementation Guidelines

Version 1.35 October 2015

5G Tech Lab, Corporate R&D Center, SK telecom

SK Telecom's 5G Architecture Design & Implementation Guidelines

Version 1.35 October 2015

5G Tech Lab., Corporate R&D Center, SK telecom

Executive Summary

Over the past few years, “5G” has definitely been one of the hottest keywords in the telecommunications as well as ICT industry. Korea revealed its ambitious plan to become the very first nation to demonstrate and commercialize 5G network with innovative services as early as 2018 and 2020 respectively. SK telecom, the leading operator in Korea, is aggressively striving towards achieving and realizing the ambition.

So far, many research organizations, industry alliances, equipment vendors, and even governments around the world have spent tremendous time and efforts to identify potential 5G services and key enabling technologies for 5G. The insights learnt during this process are absolutely necessary and indispensable. However, they are yet quite abstract and intangible.

Given the collected insights learnt so far, we believe it is the right time to organize the insights learnt so far and move towards the next step, which is transforming these insights into a more concrete and tangible form: potential 5G architecture and guidelines on how the architecture can actually be implemented.

This document presents SK telecom’s initial thoughts on 5G service types, architecture, and the guidelines on how the architecture should be implemented. In this document, readers will find SK telecom’s initial thoughts and descriptions on:

① 3 major 5G service types

② 5G architecture design, driven from the service types and their requirements

③ Implementation guidelines for realizing and platformizing the 5G architecture

We would like to emphasize that, what is presented in this document is a first step towards actual implementation of potential 5G architecture, and is not the end picture.

It is our hope that the document serves as a trigger and further stimulates the on-going 5G discussion towards a more concrete and tangible architecture and implementation.

We also hope that the constructive comments and feedbacks are continuously and iteratively collected, incorporated, and improve our final architecture for timely implementation and commercialization of 5G.

5G Tech Lab

Corporate R&D Center, SK telecom

Contents

1. Introduction

1.1. Purpose

The main purpose of this ***5G architecture design and implementation guideline*** is to present initial thoughts on SK telecom's 5G architecture, how it is designed, and how best to implement the designed architecture and establish a common understanding, internally within SK telecom as well as externally with equipment vendors and other partners.

As such, the document is to be used internally within SK telecom across relevant teams and departments for making an effective decisions and harmonized (r)evolution towards 5G.

At the same time, the document will hopefully provide external equipment vendors and other potential partners with sufficient details on SK telecom's 5G architecture design for a consistent understanding. This common understanding between SK telecom and external partners will lead to a more timely and effective implementation of the designed architecture and also jointly buliding and fostering 5G ecosystem.

Both internally and externally, a clear and consistent understanding on the architecture will lead to constructive comments and suggestions on how the architecture, its design, and the implementation methodologies can improve. It is expected that the 5G architecture described in this document continuously enhances as the constructive feedbacks collected overtime are incorporated iteratively.

The main purpose of this 5G architecture design and implementation guideline is to present initial thoughts on:

① ***SK telecom's 5G architecture***

② ***how it is designed***

③ ***how best to implement the designed architecture***

for establishing a common understanding, internally within SK telecom as well as externally with equipment vendors and other partners.

1.2. Scope

The document covers initial thoughts on SK telecom's 5G architecture, its design, and implementation guidelines on how best to build it.

We recognize that OSS, BSS, and how they are integrated with the core 5G system are important parts of 5G architecture design, especially with the trend of emergence of various B2C and B2B vertical services. However for clarity, we intentionally focus on the technical aspect of 5G system and do not include OSS/BSS and potentially other business aspect in this document.

2. Definitions, Abbreviations and References

2.1. Definitions

Term	Description

2.2. Abbreviations

Term	Description
5G	5th generation mobile networks
AAS	Active antenna system
AC	Architectural challenges
API	Application programming interface
BBU	Baseband unit
C-Core	Cloud Core
C-RAN	Cloud RAN
CAPEX	Capital expenditures
CCC	C-Core implementation considerations
CCG	C-Core implementation guideline
CDN	Content distribution network
CoMP	Coordinated multi-point
COTS	Commodity off the shelf
CPRI	Common public radio interface
CRG	C-RAN implementation guideline
DDoS	Distributed denial of service
DU	Digital unit
E2G	End-to-end 5G infrastructure implementation guideline
eICIC	Enhanced inter-cell interference coordination
eNB	e-NodeB

EPC	Evolved packet core
FBMC	Filter bank multicarrier
HR	High level requirement
HSS	Home subscriber server
IoT	Internet of things
KPI	Key performance indicator
LTE(-A)	Long term evolution (advanced)
MIMO	Multple input multiple output
MME	Mobility management entity
MTC	Machine type communication
NDN	Named data network
NFV	Network functions virtualization
NGMN	Next generation mobile networks
NOC	Network operations center
NOMA	Non-orthogonal multiple access
O&M	Operations and management
OPEX	Operational expenditures
OPNFV	Open platform for NFV
OTA	Over the air
PCRF	Policy and charging rules function
PDN	Packet data network
QoE / QoS	Quality of experience / Quality of service
RAN	Radio access network
RAT	Radio access technology
RF	Radio frequency
RU	Radio unit
SAE	System architecture evolution

SDN	Software defined networking
SDO	Standards developing organization
SCEF	Service capability exposure function
SLA	Service level agreement
ST	Service type
TCO	Total cost of ownership
VNF	Virtualized network functions

2.3. References

	Title
1	SK telecom 5G white paper http://www.sktelecom.com/img/pds/press/SKT_5G%20White%20Paper_V1.0_Eng.pdf
2	5G: New wave towards future societies in the 2020s (by 5G Forum, Korea) http://www.5gforum.org/#!/c7bg
3	5G Vision, requirements, and enabling technologies (by 5G Forum, Korea) http://www.5gforum.org/#!/c7bg
4	NGMN 5G white paper https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf
5	4G Americas' recommendations on 5G requirements and solutions http://www.4gamericas.org/files/2714/1471/2645/4G_Americas_Recommendations_on_5G_Requirements_and_Solutions_10_14_2014-FINALx.pdf
6	METIS: Final report on architecture https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D6.4_v2.pdf
7	Understanding 5G: Perspectives on future technological advancements in mobile (by GSMA) https://gsmaintelligence.com/research/?file=141208-5g.pdf&download
8	Low latency 5G architecture for mission-critical IoT (Internet of Things), The 30th International Technical Conference on Circuits/Systems, Computers and Communications, June 2015
9	To-be Architecture – Principles & Guidelines, July 2014 (SKT internal document)

3. 5G service types, requirements and challenges

Over the past few years, various research organizations, service providers, and equipment vendors have spent time and efforts to better understand and define what 5G may be and the types of services that might newly appear in 5G or be improved with 5G [1][2][3][4][5][7].

In this section, we categorize the introduced 5G services so far, based on the technical requirements that the services pose on the network. The identified service types are then used as the basis to derive SK telecom's 5G architecture and the guidelines towards realizing the architecture.

3.1. Major 3 service types (STs)

There are an overwhelming number of new 5G services that appear in various research documents and white papers. Initial studies on the new services and their requirements lead to a common conclusion that the services can be categorized into mainly 3 different service types.

The major service types are (1) virtual experience and media [ST1], (2) massive-connectivity machine type communication [ST2], and (3) mission-critical machine type communication [ST3]. The following table lists the three service types, along with example services that belong to the identified service types:

Service Type ID	Service Examples
[ST1]	• Real-time delivery of high quality content (e.g., Super multiview content, 8K UHD, etc.) delivery • Immersive tele-presence (e.g., smart home, smart office, smart city, etc.) • Massive high quality content sharing • AR/VR based content display and interaction • Super multiview content sharing and delivery • Video based life logging or surveillance (e.g., video based blackbox for bicycles and vehicles, etc.) • Broadcast services
[ST2]	• Smart metering (e.g., gas, electricity, etc.)

Various research organizations, service providers, and equipment vendors have recently described what 5G may be and the types of services that might newly appear in 5G or be improved with 5G.

The major service types are:

① ***Virtual experience and media***

② ***Massive-connectivity machine type communication***

③ ***Mission-critical machine type communication***

	• Smart environment management (e.g., smart farms, smart fishery, etc.) • Smart grid and sensor networks • Personal (and wearable) sensors
[ST3]	• Vehicle to anything (V2X) for safety enhancement • Assisted/autonomous driving & parking • Remote controlled machines • Industry 4.0 • Tactile Internet • Public safety networks • eHealth (e.g., remote health check, remote surgery, etc.)

3.1.1. Virtual experience and media [ST1]

Currently, the overall amount of multimedia contents takes a significant portion (e.g., more than 60% of overall mobile traffic) and is still growing. It is projected that this trend continues in the upcoming years, as the user mobile devices will continue to become increasingly powerful to provide users with a more vivid and immersive service experiences (e.g., 3-dimensional ultra-high definition video, augmented reality, virtual reality, etc.). This service type calls for redesign of the architecture to efficiently and scalably deliver tremendous amount of both peak and overall real-time content traffic.

__Virtual experience and media service type__ requires the underlying network and infrastructure to deliver __massive amount of traffic in real-time with minimal delay__.

Some examples services are:

- ***3-dimensional ultra-high definition video***
- ***Super multi-view video streaming***
- ***Massive content sharing over social network services***
- ***Tele- presence***
- ***Augmented reality***
- ***Virtual reality***

3.1.2. Massive-Connectivity Machine Type Communication [ST2]

As the technologies become mature and start to become available at low cost, increased number of devices and things are connected to provide rich information about individual users as well as the surroundings without human intervention. Although the data being sent and received by the individual devices belong to this category is expected to be best-effort basis and relatively small amount, this service requires the underlying future 5G communication system to be capable of handling massive number of potentially simultaneous control plane connectivity.

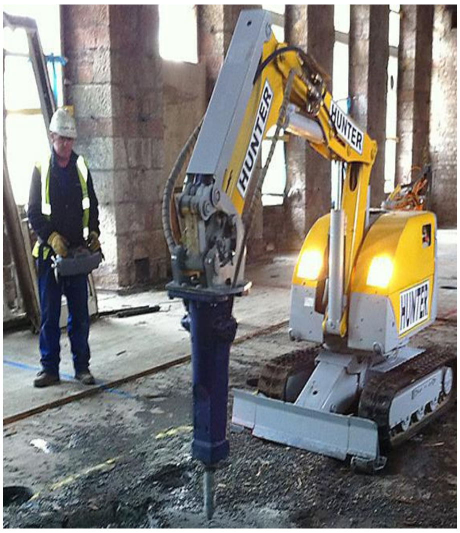
Massive-connectivity machine type communication service *requires the underlying future 5G communication system to be capable of handling* ***massive number of potentially simultaneous connectivity.***

Some example services in this service type are:

- ***Smart city, home, and life***
- ***Smart metering***
- ***Smart environment management***
- ***Smart wearables***

3.1.3. Mission-Critical Machine Type Communication [ST3]

As the technologies used to implement mobile networks and how the networks are managed for high performance, reliability, and availability are better understood over time, various mission critical services (e.g., remote machinery, vehicle, drone controlling, public safety networks, etc.) start to be implemented on top of mobile networks. The capability to send and receive data and control information wirelessly across large metro areas reliably with minimal end-to-end latency makes the future mobile network an ideal network solution, and this demand seems to only increase in 5G. These services require the underlying communication system to be highly reliable and available even in the case of large scale natural disasters.

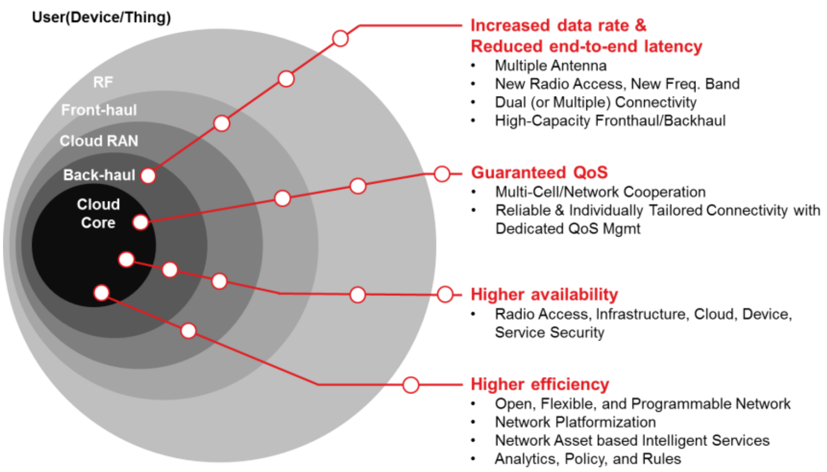


Mission-critical machine type communication service type *requires the underlying communication system to be* ***highly reliable and available*** *even in the case of large scale natural disasters.*

Some example services are:

- ***Remote machinery, vehicle, drone controlling***
- ***Smart eHealth***
- ***Public safety networks***
- ***Tactile Internet***

3.2. 5G Requirements



※ The whole 5G system is divided into 5 segments, namely Cloud Core, backhaul, Cloud RAN, fronthaul, RF, and device. Circles are placed for each requirements on the most impacted segments.

5G newly introduces many innovative services with distinguished values. These new services pose diverse requirements, which can be organized mainly into the following five requirements:

① ***Increased date rate***

② ***Reduced end-to-end latency***

③ ***Massive connectivity***

④ ***Guaranteed QoS***

⑤ ***Higher availability***

⑥ ***Higher efficiency***

3.2.1. Increased data rate

5G promises from tens to thousand times improvement in the overall data rate compared to the current 4G system, and this increase is measured from the view point of individual users as opposed to the best possible peak data rate measured from the network perspective.

5G will support several tens of Gbps to guarantee consistent user-experience, served at Gbps-level on the average anywhere and anytime. More specifically, the 5G system shall serve the followings:

- 100Mbps ~ 1Gbps per user, anywhere anytime
- 20 Gbps peak data rate

Increased bandwidth:

- ***100 Mbps ~ 1 Gbps per user, anywhere anytime***
- ***20 Gbps peak data rate***

3.2.2. Reduced end-to-end latency

The emergence of various remote controlled devices (e.g., drones, robots, machines, etc.) will eventually lead to the necessity of a highly reliable wireless communication channel with extremely low delay. Growing demands for autonomous cars are another motivation for such ultra reliable and low latency wireless communication specified in IMT-2020.

More specifically, 5G system shall support the followings:

- 1 ms radio latency (one way)
- 10 ms (or less) end-to-end latency
  (e.g., one way from UE1 to UE2)

Reduced end-to-end latency:

- *1 ms radio latency*
- *10 ms (or less) end-to-end latency capability*

3.2.3. Massive connectivity

Increased number of devices and things (e.g., sensors, meters, etc.) is now being connected, and the number of connected devices is expected to increase at a faster rate. Although the amount of data being communicated by these devices and things in the data-plane is relatively smaller than high quality multimedia contents, they tend to trigger frequent control-plane messages in bursts simultaneously.

More specifically, the following connectivity density shall be provided:

- 10^6 connections per km^2

Massive connectivity:

- *10^6 connections per km^2*

Many of the use cases requiring massive connectivity operate in best-efforts basis. Thus, the actual number of devices and things may be much more than specified above. In such cases, the service shall be isolated from other services (or infrastructures) such that the service congestion does not affect quality of any other services being provided.

3.2.4. Guaranteed Quality of Experience

The massive number of devices connected to the network leads to very rich set of information (e.g., time, location, and context) for the operator and other 3rd party players, which may be used to provide personalized services. Operators and 3rd party players will use the information to best utilize and enhance quality of experience, taking the full advantage of dynamic and programmable network infrastructure and real-time data analytics.

At the same time, a subset of gathered information will potentially be exposed to the 3rd party application and service providers, such that the overall service performance may be optimized based on the information provided by the network (e.g., network load, and congestions).

3.2.5. Higher availability

One of the representative 5G use-cases is mission-critical MTC (e.g., public safety network). More specifically, 5G is designed to be highly available and reliable with zero-perceived service downtime

5G services pose a number of qualitative requirements *such as* to support various mission critical services. The main difference compared to 4G would be cloud-friendliness, as 5G will mostly be virtualized and take full advantage of built-in HA/DR features of cloud. Ultimately, the goal is to provide zero-perceived failure time from the end-to-end service perspective.

***guaranteed quality of experience** and **high availability** (e.g., zero-perceived failure time), which may be later assessed by the service users individually.*

This requirement can be achieved by having in-network analytics, intelligence along with end-to-end orchestration.

Furthermore, 5G network is expected to provide higher security and robustness, compared to the existing 4G system.

3.2.6. Higher efficiency

One of the key enablers for high throughput in 5G is densely-deployed small cells that lead to higher CAPEX and OPEX. In order to take full advantage of this massive-scale small cells, 5G network shall be a lot more efficient in cost and energy usage, given that the massive deployment of infrastructure, devices, and the overall cost should be efficiently managed in a scalable manner.

Higher efficiency:

- ***100x greater energy efficiency than 4G***
- ***10 Mbps per m^2***

From the view point of mobile operators, the reduction in the total cost of ownership (TCO) is especially important. More specifically, the following efficiency in energy and traffic delivery requirement shall be met.

- 100 times greater energy efficiency than that of 4G
- 10 Mbps per m^2

For the reference, the following table lists the 5G architectural requirements (vs. 4G requirements) that must be met, defined by ITU-R WP5D.

	5G (IMT-2020)	4G (IMT-Advanced)
Peak data rate	20 Gbps	1 Gbps
User experience data rate	100 Mbps ~ 1 Gbps	10 Mbps
Spectral efficiency	3x greater than 4G	
Mobility	500 km/h	350 km/h
Latency	1 ms (radio)	10 ms (radio)
Connection density	10^6/km^2	10^5/km^2
Energy efficiency	100x greater than 4G	
Area traffic	10 Mbps/m^2	0.1 Mbps/m^2

***5G is user-centric**, as opposed to being network-centric in the past, **and disruptive in nature**, integrating different industries to create **enabling services with enhanced overall service experience** of the end-users.*

To summarize, 5G is user-centric, as opposed to being network-centric in the past, and disruptive in nature, integrating different industries to create enabling services with enhanced overall service experience of the end-users.

3.3. Architectural challenges

There are architectural challenges associated with each of the architecture requirements identified in the previous section.

The exisitng architecture has several architectural challenges when considering the requirements elaborated in the previous section.

In this section, we list the corresponding architectural challenges for each of the requirements discussed.

3.3.1. Increased data rate

As it is practically infeasible to achieve such a high data rate based on how today's infrastructure is designed, it is suggested that ***a new radio carrier at a much higher frequency band (e.g., above 6 GHz) to be considered****.*

In addition to the radio access network challenges, ***the core network also needs to be re-architected to become flatter and distributed to remove any potential traffic bottlenecks****.*

Increased amount of bandwidth poses architectural challenges on both radio access and core network segments. So far, one of viable approaches to increase the peak data rate has been aggregating multiple carriers together and uses them simultaneously to send or receive data which belong to a single flow. The peak data rate of LTE-A commercial services by SK Telecom is 450Mbps at the time that this document is written, achieved by aggregating three different carriers and 2x2 MIMO.

Note that this is the best-case peak data rate, which is very different from Gbps data rate required for 5G. Because it is practically infeasible to achieve such a high data rate based on availability of frequency bandwidth, it is suggested that a new carrier frequency above 6 GHz where we can find higher bandwidth should be considered.

In addition to the radio access network challenges, the core network also needs to be re-designed and optimized (e.g., remove any potential traffic bottlenecks) because the LTE-SAE network nodes are currently structured in a hierarchical manner, where all user traffic traverse through a series of serving gateways (S-GWs) and packet data network gateways (P-GWs). In order to handle the data-plane traffic more efficiently, the network shall become flatter and more distributed to avoid the user traffic from unnecessarily reaching the network nodes located remotely in the center of the network.

3.3.2. Reduced end-to-end latency

To minimize the overall delay (e.g., to be less than 10ms), ***the entire system should be redesigned and optimized with a holistic view*** *in mind.*

Today with LTE, it typically takes about several tens of milliseconds for a given IP packet sent from an application running inside a device to traverse up and down the LTE protocol stack before finally received by the receiving application. This latency is mostly due to how LTE (e.g., Transmission Time Interval and radio access protocol, etc.) is designed and packet routings in mobile backhaul networks.

To minimize the overall end-to-end latency (e.g., to be less than 10ms), the entire network infrastructure and the protocols should be redesigned and optimized with a holistic view in mind. These system redesign must be done efficiently with care, as all of these architecture redesign and optimization can potentially increase the overall cost of network functions, infrastructure, and management significantly. Initial observations on how to reduce latency from end-to-end perspective can be found in [8].

3.3.3. Massive connectivity

__The control-plane of the future mobile system shall be optimized__ to efficiently handle such traffic generated by machine type communication.

The existing LTE-SAE mobile networks are designed and optimized mostly for data-plane traffic. That is, the system is more efficient in handling data-plane traffic than control-plane traffic.

Recently and in the upcoming years, different traffic pattern is and will continue to be observed for machine type communication. These devices and things make very frequent and sporadic attempts to transmit small amount of data, leading to huge amount of overall and simultaneous control-plane signaling traffic and small amount of overall user plane traffic.

To avoid having a potential network collapse caused by the massive simultaneous control-plane signaling traffic, the control-plane of the 5G core networks shall be optimized to efficiently handle such traffic generated by machine type communication.

3.3.4. Guaranteed Quality of Experience

Taking the full advantage of the __flexible and programmable network__, it becomes feasible to __configure and optimize the network performance on-demand in real-time__.

In addition, __virtualizing and decoupling core software functions__ from the underlying h/w __leads to high availability__.

__A practical question__ on how these solutions can be applied to build the

Recently, there have been a lot of efforts to make the mobile networks flexible and programmable. At the same time, techniques on collecting huge amount of network data and analyzing the collected data in real-time have also been maturing. Taking full advantage of the flexible and programmable networks, it becomes feasible to configure and optimize the network performance on-demand in real-time.

Some of the examples are ETSI NFV Management and Orchestration (MANO) and NETCONF-based dynamic network configuration automation.

3.3.5. Higher availability

Virtualizing and decomposing core software functions from the underlying hardware and cloudified environment yields several benefits. One of the main benefits is high availability where the function can be migrated and replicated flexibly.

Another aspect for higher availability is security. Although the current methodologies and procedures for security will still be important, a new approach (such as zero-trust security) to ensure high security will be needed as the current premise-based security will no longer be able to guarantee security in the network becomes cloudified.

overall 5G system in a cost-efficient manner ***still remains*** *and is on-going.*

3.3.6. Higher efficiency

5G network is expected to provide a higher data rate at any place, which eventually leads to a higher overall deployment cost. Furthermore, 5G services and use-cases pose extremely stringent technical requirements, which may potentially raise the cost of individual network functions to be deployed. Therefore, individual network functions and the overall infrastructure have to be highly cost-efficient. Fortunately, the industry has diligently worked on addressing each of the above challenges individually. On the other hand, a practical question on how these solutions can be applied to make up the overall 5G system in a cost-efficient manner still remains; applying the related and inter-dependent technologies altogether on a single 5G system may potentially lead to an overly complex and cost-prohibitive system.

4. 5G architecture design and implementation guidelines

4.1. General description of SK telecom's 5G architecture

The following 5G architecture is designed, based on the services and their requirements described in the previous sections.

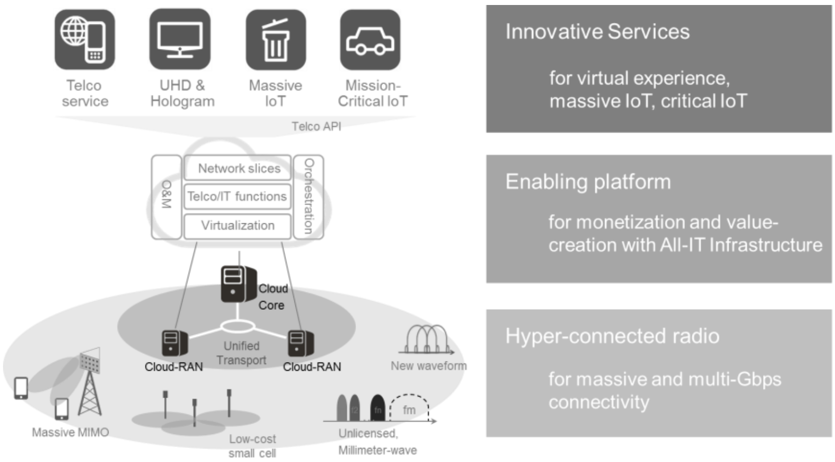
The 5G architecture is mainly composed of three horizontal layers, namely "innovative service enablement", "enabling platform", and "hyper-connected radio", which we describe in this section in detail.

SK telecom's 5G architecture consists of the following three layers:

① ***Innovative service enablement layer*** *includes innovative 5G services*

② ***Enabling platform layer*** *transforms the mechanical and disjointed underlying networks to a service-oriented infrastructure*

③ ***Hyper-connected layer*** *can deliver massive amount of data to the "Enabling Platform" layer in a very efficient and seamless manner.*



① **Innovative service** is the top layer that includes innovative 5G services. Built upon differentiated 5G infrastructure capabilities, 5G services will be able to offer a highly reliable and immersive user experience. In particular, ultra-high definition video streaming will eventually become dominating services that provide virtual user experiences such as multi-view video streaming and even holograms. Furthermore, the sufficient bandwidth and high reliability offered by 5G will facilitate the emergence of remote robot controlling services and mission-critical Internet of Things (IoT) services.

② **Enabling Platform** is the middle layer that creates meaningful and service-centric values to support the upper "innovative service" layer by properly transforming the mechanical and disjointed underlying networks to a service-oriented infrastructure. It utilizes two key enablers to perform this task. The one is Network Functions Virtualization (NFV) and Software Defined Networking (SDN), which makes the network much more dynamic, agile, on-demand, and flexible. The other is a well-defined set of Application Programming Interfaces (APIs), which offers the ability to automate and orchestrate the network both internally for automated network.

③ **Hyper-Connected radio** is the bottom layer that can deliver massive amount of data to the “Enabling Platform” layer in a very efficient and seamless manner. Due to the massive amount of transmission data and stringent latency requirements, this layer will include the not only the existing radio network technologies (both as-is and evolved form), but also new novel radio network technologies. It is necessary to develop a range of technologies capable of enhancing cell splitting, spectral efficiency, channel frequency bandwidth and network operation efficiency.

While both 3G and 4G radio access networks (RANs) were built as a stand-alone network, 5G RAN will be deployed by integrating the existing LTE-Advanced (LTE-A), its evolution technologies, and new radio access technologies (RATs). Due to their heterogeneous nature, it is important to build an infrastructure where different radio access technologies are seamlessly integrated.

5G RAN will be deployed by integrating the existing LTE-Advanced (LTE-A), its evolution technologies, and new radio access technologies (RATs).

3GPP has been developing standards for LTE-A evolution in Release 13 and plans to start standardizing the new RAT in Release 14 from 2016. Therefore it is important to continuously develop both technologies to realize 5G RANs

4.2. End-to-end 5G infrastructure

4.2.1. Description

We believe there will eventually be two types of clouds in 5G, which will be used to build the high level architecture illustrated in the previous section:

- Cloud radio-access networks (Cloud-RAN)
- Cloud core networks (Cloud-Core)

With NFV, it is anticipated that most (if not all) mobility network functions will be virtualized for flexibility and efficiency. With SDN, the control plane functions will further be separated from the data plane functions for consistency and agility.

5G system will eventually be based on mainly two types of clouds:

① ***Cloud radio-access networks (Cloud-RAN)***

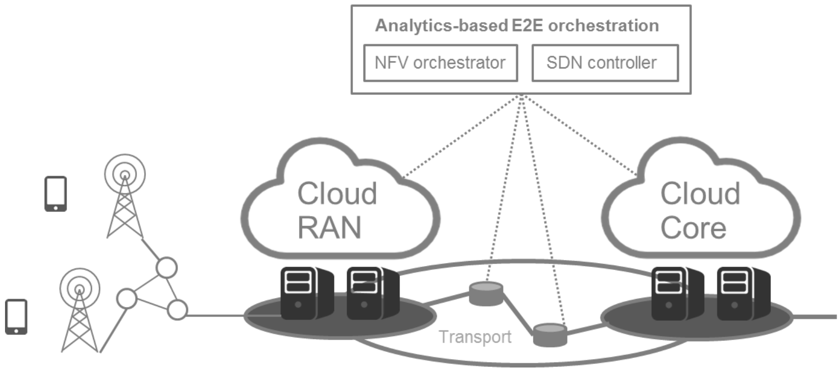
and

② ***Cloud core networks (Cloud-Core).***

The virtualized and de-coupled functions will then be placed and run appropriately in the two cloud types.

It is emphasized that, to intelligently manage QoS (quality of service) and optimize latency dynamically in real-time as the network condition changes, end-to-end network service orchestration of both network functions and network is essential.

Furthermore, SDN controller will also play an important role to enable centralized control of routers in transport networks between the two Cloud types.



4.2.2. E2E 5G infrastructure implementation guidelines (E2Gs)

End-to-end 5G infrastructure design and guidelines are derived from the view point of all-IT and associated IT technologies.

Here, overall guidelines on how the 5G infrastructure can be practically implemented are described. These guidelines are first derived from the viewpoint of all-IT virtualized infrastructure [9], virtualized resource abstraction and pooling, orchestration, and security. Then the derived guidelines are further organized into 5 different aspects, namely implementation (IMP), deployment (DEP), security (SEC), operation & management (O&M), and additional function (ADD) aspects. Within each aspects, the guidelines are ordered based on the generality (e.g., general → specific).

The following table summarizes the E2E 5G infrastructure requirements, organized into the 5 aspects and how they address the 5G requirements described earlier in this document.

Aspect	*ID*	*D*	*L*	*C*	*Q*	*A*	*E*	*Guidelines*
IMP	E2G.1			○		○	○	Use of industry standard and general-purpose H/W
	E2G.2			○			○	Use of standard networking nodes / protocols

Requirements

D: Increased Data Rate
L: Reduced Latency
C: Massive Connectivity
Q: Guaranteed QoE
A: Higher Availability
E: Higher Efficiency

	E2G.3	○	○		○			Carrier-grade performance and reliability
	E2G.4				○	○		Resilient fronthaul & backhaul
	E2G.5	○	○	○	○	○	○	Support 5G KPIs specified by ITU-R WP5D and potentially other indusrial organizations such as NGMN
DEP	E2G.6		○	○			○	Distributed infrastructure and unified end-to-end orchestrator
	E2G.7						○	Use of optical fiber infrastructure
SEC	E2G.8					○		Protected access and control to infrastructure resources
	E2G.9					○		Support per-customer data encryption and integrity assurrance
	E2G.10					○		Immunity to virus software and external intrusion
	E2G.11				○	○		Support for zero-trust security
O&M	E2G.12				○	○	○	Support of means and tools for efficient infrastructure management
	E2G.13				○	○	○	Support of unified control by orchestrator(s)
	E2G.14						○	Open API support
ADD	E2G.15						○	Support for independent evolvability
	E2G.16					○	○	Support for partial replacement and incremental upgrades for scalability with minimized sunken cost
	E2G.17						○	Green and environment-friendly hardware and technology

4.2.2.1. Implementation aspect (IMP)

[E2G.1] Use of standardized and general-purpose H/W

- 5G infra shall be based on industry standard commodity-off-the-shelf (COTS) IT / datacenter hardware for equipment vendor independency
- 5G Infra shall be based on hardware that allows abstraction such as virtualization / pooling & distributed computing using virtualized resources. This in turn leads to stardardized modeling of the physical infrastructure (e.g., via YANG)

5G Infrastructure shall be based on all-IT COTS. The physical resources will then be virtualized and managed as virtualized resources.

- Information on physical resource such as type of hardware, function, performance indicator shall be available to the upper layer entities such as virtualized infrastructure manager and orchestrator

- Suppliers shall provide both qualitiave and quantitative analysis of CAPEX/OPEX for COTS hardware, compared to those of vendor-specfic hardware

- Suppliers shall provide compatibilities of all performance acceleration capabilities & features and quantified performance analysis of the COTS hardware being provided.

- Performance acceleration technologies shall be (open) standard-based

- Seamless partial replacement and incremental upgrades of the infrastructure shall be supported. This incremental upgrades shall not impact service continuity

[E2G.2] Use of standard networking protocols (3GPP, IETF)

The connectivity amongst the network functions and how they are orchestrated shall be standard-based.

IP is the basic protocol used for inter-node communication. When needed, standard-based IP overlays and tunnelling should be supported

- 5G infra shall provide IP connectivity at minimum

- 5G infra shall provide dynamic interconnection of multi-networks with full capability of orchestration and management

- The interconnection shall support differentiated QoS management at both the device level (e.g., NICs), host and network (e.g., links and paths) level

- 5G Infra should be designed with minimized physical and virtual (e.g., overay/underlay network topology) hierarchies

- Control plane and data plane shoud be separated for higher efficiency and scalability of 5G infra. The communication between control / data plane should be (open) standard-based.

[E2G.3] Carrier-grade performance and reliability

All network functions shall be designed and optimized with carrier

- (Open) standard-based hardware accelaration features should be supported and offered to the upper layer

- Virtualized network functions must be redesigned to be cloud-friendly, e.g., to take full advantage of built-in HA and DR capabilities of virtualized environment. That is, service continuity shall be guaranteed through resource abstraction layer, and the virtualized network functions shall be designed to take full advantage of the HA/DR capability of the resource abstraction layer

- All failures of physical and virtual elements shall be monitored in real-time

- Seamless lifecycle management of virtualized application shall be supported with minimal overhead

- Guaranteed isolation between virtualized resources and functions for protection and security

grade reliability and performance.

In other words, virtualization of network functions or network itself shall not result in performance degradation.

[E2G.4] Fronthaul and backhaul shall be designed to be resilient to failures

[E2G.5] 5G RAN, Core, and Transport must support 5G KPIs specified by ITU-R WP5D and potentially other industrial organizations such as NGMN

4.2.2.2. Deployment aspect (DEP)

[E2G.6] Distributed infrastructure for performance

- Edge cloud shall support both radio functions as well as distributed core functions to reduce backhaul traffic and for performance enhancements

- Edge cloud may be co-located in C-RAN where virtualization technology is used to implement radio and other functions on unified hardware platform.

- The interconnection between the clouds shall support differentiated quality of service in a stardardized manner (e.g., segment routing) to better support virtual network isolation and protections (e.g., in case of network slicing)

The infrastructure should be distributed efficiently interconnected for more efficient traffic handling at the edge.

[E2G.7] Optical fiber infrastructure

- Additional deployment of optical fiber infrastructure

Additional deployment of optical fiber infrastructure shall be shall be minimized

minimized.

Transport networks shall also be flexible and monitored by central orchestrator.

- Switch/controllers newly deployed or replaced in transport networks shall support southbound/northbound API interface fully compliant to SDO, e.g. Transport SDN in ONF
- The use of optical wavelength in transport network is compliant to ITU-T fixed grid or be based on elastic optical networking that allows the use of flexible grid only under the condition that the usage information of wavelength is provided through northbound API
- It shall be considered for the optical routers to support lower layer switching where needed

4.2.2.3. Security aspect (SEC)

[E2G.8] Protected access and control to infrastructure resources

As the overall system becomes more virtualized, open, distributed and large in scale, the system shall be robust and secure.

Security models which better support virtualized environment such as zero-trust security shall be supported.

- Resource abstraction layer shall provide full capability of virtualization, pooling/distributed computing transparently
- All applications shall not be able to directly access physical resource including sever, storage and networks
- Advanced cloud features such as live migration, scaling, and healing shall be capable independently from S/W through resource abstraction layer

[E2G.9] Support per-customer encryption

It is important that the operations and management of physical infrastructure to be completely isolated for consistency and security.

- Use and management of encryption shall be applied for customer and confidential information
- The entities that encryption are applied for shall be timely updated according to goverment regulation and relevant laws

[E2G.10] Immunity to external intrusion and viruses

- Detection and protection of external intrusion shall be provided for the overall network, especially for the connection point to external networks

- Anti-virus software shall be intalled to secure systems from DDoS and malicious code attack.

[E2G.11] Support for zero-trust security

- Premise(zone)-based security shall be honored as a fundamental security measure
- Per-VM security, as the physical premise(zone)-based security no longer sufficient in the virtualized environment
- Relevant security network functions (e.g., firewalls, access control nodes, etc.) shall be per-VM granularity if needed

4.2.2.4. Operation and management aspect (O&M)

[E2G.12] Support of means and mature tools to efficiently manage and monitor the infrastructure

- Suppliers shall provide mature tools and practices to manage and monitor the infrastructure

[E2G.13] Support of unified control by orchestrator(s)

- Policy and SLA-based virtual resource management shall be supported via orchestrator
- Live migration across different datacenters and COTS server groups shall be supported. If needed (for performance reasons where the virtual servers cannot satisfy the required SLA), live migration to even physical servers shall be supported.
- Support of seamless inter-DC resource pooling and failover
- Dynamic routing and interconnection between virtual overaly network and physcial network shall be supported
- Zero service down-time shall be guranteed during S/W upgrade and change via In-Service-Upgrades (or Updates)
- Dynamic lifecycle management of a given intra/inter-

End-to-end orchestration shall support features to seamlessly lifecycle manage both the physical and virtual network functions and networks.

Furthermore, seamless interworking between the virtualized network functions and physical network functions shall be supported.

DC network function or service (such as scale-in/out, auto-healing, upgrades) shall not impact service continuity

- Network configuration for both initial network provisioning as well as run-time configuration modification shall be supported
- Support of (anti-)affinty of compute, network, storage, and physical location when needed
- Dynamic configuration changes on inter-DC bandwidth, connectivity, routing, and QoS

[E2G.14] Open API support

5G network shall have capability to provide services via open APIs to external parties, when desired by the operator.

- Both Northbound APIs (for application development) and Southbound APIs (for the control of network function, resources) shall be supported for network-as-a-service support
- Statistics of all API usage patterns shall be available for marketing and service development purposes
- The offered APIs shall be provided in secure and robust manner

4.2.2.5. Additional functions aspect (ADD)

[E2G.15] Support for independent evolvability

The core network functions shall be able to evolve independently without impacting other technology domains (e.g., radio).

- The core network functions shall be designed with minimal dependency to other technology domains (e.g., radio access) such that it can evolve indepently without impacting other technology domains.

[E2G.16] Support for partial replacement and/or incremental upgrade for scalability with minimized sunken costs

As the infrastructure becomes massive in scale, it becomes important that a part of the infrastructure can be incrementally upgradable, without being wasted and replaced completely.

[E2G.17] Green and environment-friendly hardware and technologies

4.2.3. Other E2E infrastructure implementation considerations (E2Cs)

Other implementation considerations will be added in the future.

4.3. Cloud RAN (C-RAN)

4.3.1. Description

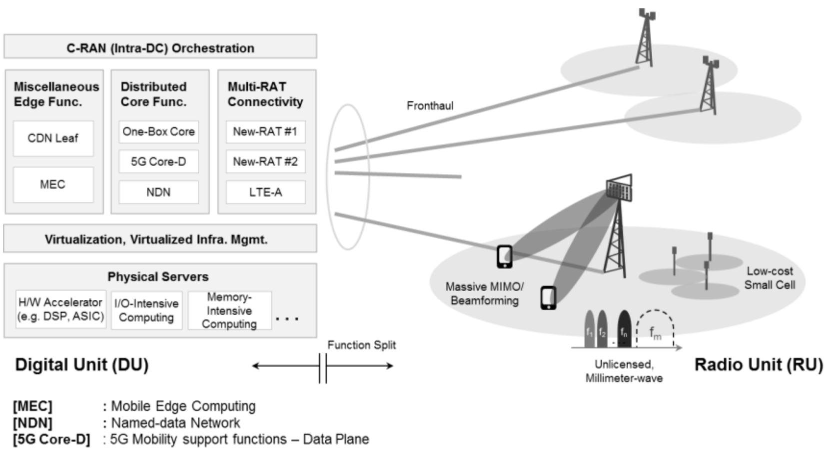
Cloud-RAN (C-RAN in abbreviation) aims to transform the radio access networks (RAN) to be flexible, scalable and service-agile by taking advantage of evolving IT virtualization and cloud computing technologies.

Not only does C-RAN serve as flexible aggregation points for better performance optimization, it can be also a key enabler for a variety of radio access network technologies, e.g., small cell deployment, new radio access technology (RAT), front-haul, and coordination among multi RAT to satisfy above-described 5G requirements.

The concept of C-RAN of 5G from our prospective is regarded as connotation of the whole access network including not only virtualized architecture but also new radio feature for 5G and front-haul between C-RAN digital unit (DU) and radio unit (RU).

This section elaborates guidelines towards practical C-RAN implementation.

***Cloud-RAN** (C-RAN in abbreviation) on 5G network aims to **transform the radio access networks (RAN) to be flexible, scalable and service-agile** by taking advantage of evolving IT virtualization and cloud computing technologies.*



4.3.2. C-RAN implementation guidelines (CRGs)

The C-RAN guidelines are categorized into the following 5 aspects: implementation (IMP), deployment (DEP), security (SEC), operation & maintenance (O&M), and additional function (ADD) aspects.

The guidelines are summarized in the following table.

Aspect	*ID*	*D*	*L*	*C*	*Q*	*A*	*E*	*Guidelines*
IMP	CRG.1	○	○	○		○	○	Comparable or higher competiveness in processing performance per unit price of industry standard IT servers,

Requirements

D: Increased Data Rate
L: Reduced Latency
C: Massive Connectivity

								compared to legacy function-specific specialized hardware
	CRG.2				○		○	Open software environment under operator’s agreement
	CRG.3			○		○	○	Standard-compliant multi-RAT defined in SDO
	CRG.4			○		○		Support of backward/forward compatibility with/of legacy network
	CRG.5			○		○		No performance degradation when interworking with legacy network equipments (e.g., eNB)
	CRG.6	○	○	○				Front-haul fully supporting connectivity function
	CRG.7	○	○	○	○	○	○	C-RAN shall support 5G KPIs specified in ITU-R WP5D and potentially in other industrial organizations such as NGMN
DEP	CRG.8			○			○	Comparable or smaller hardware size for cell-site colocation
	CRG.9						○	Comparable or lower overall OPEX
	CRG.10				○		○	Open front-haul under operator’s permission
	CRG.11					○	○	Plug & Play small cell
	CRG.12	○	○				○	Comparable RU complexity even for massive MIMO
O&M	CRG.13				○	○		Open & real-time APIs
ADD	CRG.14			○			○	C-RAN aware core functions
	CRG.15			○			○	Service-aware edge functions

Q: Guaranteed QoE
A: Higher Availability
E: Higher Efficiency

4.3.2.1. Implementation aspect (IMP)

[CRG.1] Comparable or higher competitiveness in processing performance per unit price, compared to legacy function-specific specialized hardware

C-RAN hardware shall be based on general purpose processors. However, their overall performance per cost should not be less than proprietary and specialized hardware.

- C-RAN hardware shall be based on general purpose processors
- C-RAN hardware shall provide similar or even higher competiveness compared to legacy function-specific

specialized hardware in processing performance per unit price

- Compared to legacy function-specific hardware, C-RAN harware shall provide a similar or even lower operation cost for same processing performance

- C-RAN shoud be able to split it into two main parts, digital unit (DU) and radio unit (RU) between which optical (or wireless) front-haul provides interconnection

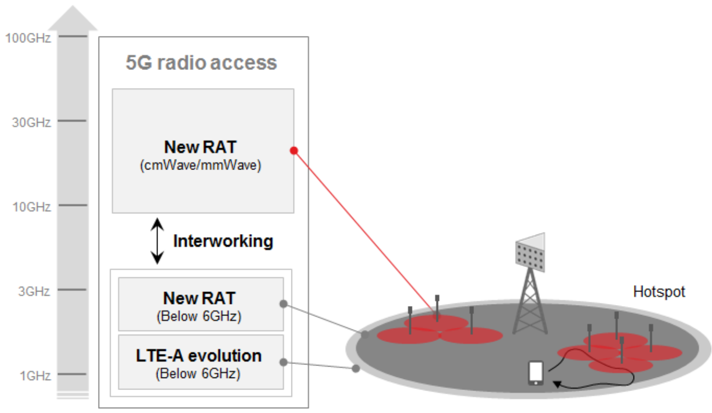
[CRG.2] Open software environment under operator's agreement

C-RAN is programmable and offers interfaces externally for 3rd party application providers.

- C-RAN shall provide programmable S/W environment where any 3rd party vendor can implement additional features and functionalities in software

- Secure administration tool to keep governance shall be provided for C-RAN

- Both northbound APIs (for application development) and southbound APIs (for the control of network function, resources) shall be supported



[CRG.3] Standard-compliant multi-RAT defined in SDO, e.g., 3GPP, IEEE

The air interface shall be based on standardized protocols.

- C-RAN connectivity function shall be fully compliant to a standard such as 3GPP, IEEE and shall also provide multi-RAT capability (e.g., LTE, new RAT below 6GHz, new RAT above 6GHz) including seamless interworking and QoS management

- C-RAN should be able to accomodate cellular IoT networks specified in 3GPP and other LPWA (Low power wide area) networks
- All connectivity functions shall be implemeted in software manner and be accessible by 3rd party under operator's agreement.
- Some of C-RAN connectivity functions that need faster processing to meet capacity and latency requirements may be realized with hardware accelerator.
- Customized features may be provided in C-RAN as long as they have no impact on standard features
- New RAT should promise TTI (transmission time interval) to be less than 1ms, possibly less than 0.2ms for low latency transmission

[CRG.4] Support of backward/forward compatibility with/of legacy network

C-RAN must support seamless and efficient backward / forward compatibility with legacy network.

- C-RAN shall provide full compatibility to existing radio access networks including 3G and 4G
- Seamless interworking and handover between 5G and a legacy networks shall be provided in C-RAN.
- Interface between/among different network equipments shall be standardized and open to operators and other manufacturers
- C-RAN connectivity function shall provide open interfaces among different network equipments following specification in SDO, e.g., X2 inferface for CoMP and eICIC.

[CRG.5] No performance degradation when interworking with legacy network equipments, e.g., legacy eNB hardware

The 5G C-RAN infrastructure shall support co-existence with the legacy network.

- C-RAN connectivity functions shall have no performance degradations in KPI statistics when coexisting with legacy network equipment

[CRG.6] Front-haul fully supporting connectivty function (air-interface)

- Front-haul should be able to deliever generated baseband signals from BBU to RU (vice versa) without signal quality degradation

- Front-haul should allows DU function splits (where L1 or L2 can be located in RU) and relevant connectivity functions (such as dual connectivity and CA)

[CRG.7] C-RAN shall support 5G KPIs specified in ITU-R WP5D and potentially from other industrial organizations such as NGMN

4.3.2.2. Deployment aspect (DEP)

[CRG.8] Comparable or smaller hardware size for cell-site colocation

Each component for C-RAN shall be equal or smaller physically as well as in total cost than their counterparts in 4G.

- C-RAN Baseband Unit (BBU)/Radio Unit (RU) H/W shall be similar or smaller size than legacy H/W for colocation within existing cell sites. The reduced overall size is important in reducing the site maintenance cost (including space rental cost).

[CRG.9] Comparable or lower overall OPEX

- C-RAN Baseband Unit (BBU)/Radio Unit (RU) shall be relatively more efficient in power consumption, compared to that of legacy network equipment.

- Operation failure shall be minimized with lower level than legacy networks, and shall also monitored by NOC all the time

[CRG.10] Open front-haul under operator's permission

When desired, C-RAN shall be able to provide open front-haul protocol information over CPRI.

- C-RAN shall provide open all front-haul protocol information e.g., I/Q mapper information, management field information over CPRI.

- Additional information and its protocol (e.g. phase-shiter control information for analog beamforming) may also be specified and applied ONLY under the condition that they are fully disclosed to a operator.

- Open front-haul should be able to support ring architecture topology and switching for higher resiliency

[CRG.11] Plug & Play small cell

- C-RAN should offer suitable small cell systems that can be simply deployed in plug-and-play manner without any additional maintenance & operation
- Additional interference in cell edge caused by small cell deployment shall be suppressed.

For massive and dense cell deployment, small cell shall be deployed in plug-n-play manner.

[CRG.12] Comparable RU complexity even for massive MIMO

- C-RAN shall provide a similar H/W complexity of RU regardless of the number of antenna which are mainly used for massive MIMO and beamforming.
- The complexity of interconnection between Radio Unit (RU) and antenna should be a similar to one in legacy RU with 2 X 2 MIMO system.
- RU and antenna may be integrated for active antenna systems(AAS) ONLY under the condition that C-RAN provides OTA measurement of RF signals
- Antenna form factors should be minimized with more eco-friendly shape and material

Hardware complexity for Massive MIMO RU shall be minimized.

More specifically, its complexity shall be similar to the legacy RU.

4.3.2.3. Security aspect (O&M)

N/A

4.3.2.4. Operation & Maintenance aspect (O&M)

[CRG.13] Open & real time API

- C-RAN shall provide a control/data information in operator-define format for centralized O&M.
- Upon the request from operators, some of control/data information shall be provided by every second without any impact on other features

C-RAN in general shall be programmable. When needed, this programmability needs to support real-time capability.

4.3.2.5. Additional function aspect (ADD)

[CRG.14] C-RAN aware core functions

- C-RAN should be able to run flexible distributed core functions for specific purposes, e.g., Enterprise EPC (or one-box EPC) for private network.

[CRG.15] Service-aware edge functions

- C-RAN should be able to provide service-aware edge cloud functions for new business and customer experience enhancement.
- These service-aware edge functions are expected to offer mainly the following three benefits:
  - Reduction in end-to-end latency
  - Reduction in data plane traffic that are unnecessarily sent to the core network
  - Reduction in control plane traffic for massive IoT

4.3.3. Other C-RAN implementation considerations (CRCs)

Other implementation considerations for C-RAN include:

- ***Round trip time***
- ***Timing asymmetry***
- ***Timing accuracy***
- ***Time variation***

- Front-haul should satisfy the following considerations (Tentative):
  - Round trip time (front-haul latency)

The round trip time between DU and RU is required to be less than **500μs** including fibre (cable) propagation time or any latency consumed by optical transport equipment.

- Timing asymmetry

Front-haul latency for upstream and downstream should be symmetrical. An asymmetrical trip delay between upstream and downstream front-haul transmission medium should impact the accuracy of the timing calculation between UE and DU. The front-haul transmission medium time difference shall not exceed **65ns** with a preference to be strictly less than this value.

- Timing accuracy

The accuracy of the measurement of round trip delay on the transmission medium shall meet the following requirement **+/- 16.276ns**. Between two periodic measurements of the round trip delay by BBU, the variation of the transmission medium must not exceed +/- 16.276ns with a preference to be strictly less than this value.

- Time variation

Maximum contribution jitter and wander from front-haul link to the DU frequency accuracy budget must not exceed **+/- 2 ppb** ($2x10^{-9}$). The RU clock shall be traceable to DU clock. This implies that the front-haul link must not introduce any constant frequency error. The link jitter and wander for the front-haul system will be specified with concrete values which will be defined in further study. Some promising techniques to reduce front-haul data rate and to improve RU power efficiency may only be applied after experimental verification that needs to be checked by operators

4.4. Cloud Core (C-Core)

4.4.1. Description

In LTE-SAE, all data plane traffic must go through a single node type called Packet Data Network Gateway (P-GW). This centralized and hierarchical architecture is advantageous in terms of operation and management. Yet, it could potentially pose a severe limitation when there is a huge amount of backhaul traffics in 5G.

The following figure illustrates the existing LTE-SAE architecture and the potential issues when there is a significant amount of traffic increase.

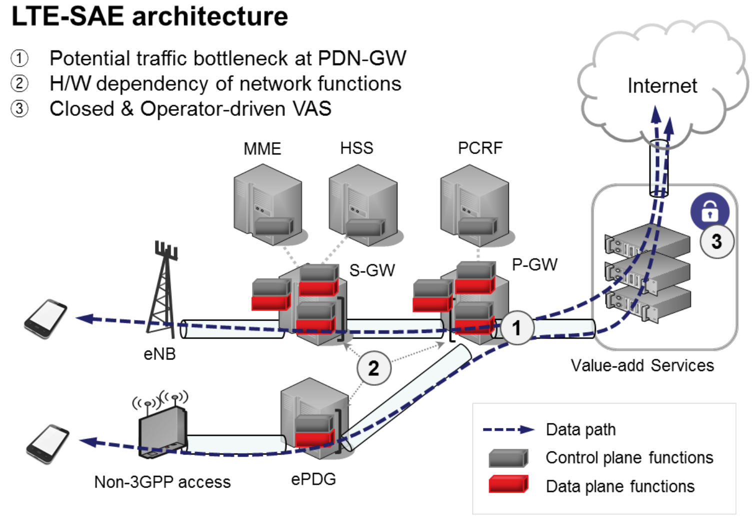
The centralized and hierarchical architecture is advantageous in terms of operation and management.

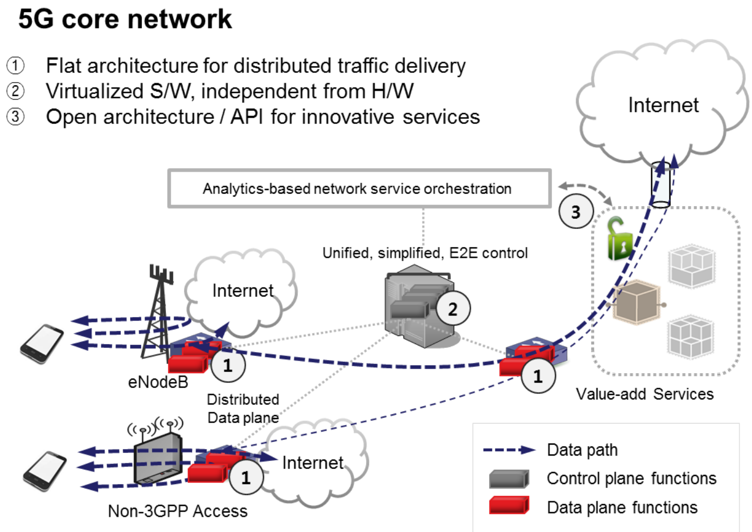
Yet, it could potentially pose a severe limitation when there is a huge amount of backhaul traffics in 5G.

Some of the engineering trade-offs and limitations of today's LTE-SAE:

① ***Potential traffic bottleneck at P-GW***

② ***H/W dependency of network functions***





At present, maximum cell capacity for 5G base station is expected to be higher than several tens of Gbps. Given that many base stations will be densely deployed and centralized for Cloud-RAN, mobile backhaul traffic are expected to exceed even more than several hundred of Gbps. Therefore, the core network functions and the overall architecture must be enhanced to better support various 5G services (such as mobile broadband, mission critical IoT, and massive IoT) even when the network is congested due to frequent traffic surges.

In order to enhance the existing architecture to better support scalability and performance, we define the following Cloud Core (C-Core) architecture and its components below.

③ ***Closed & operator-driven VAS***

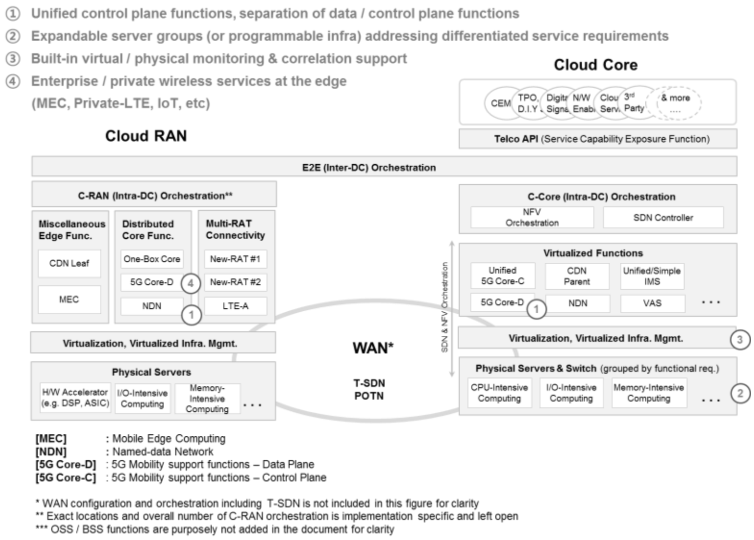
The core network functions and the overall architecture must be enhanced to better support various 5G services (such as mobile broadband, mission critical IoT, and massive IoT) even when the network is congested due to frequent traffic surges.

Some of the architectural enhancements are:

① ***Flat architecture for distributed traffic delivery***

② ***Virtualized S/W, independent from H/W***

③ ***Open architecture / API for innovative services***

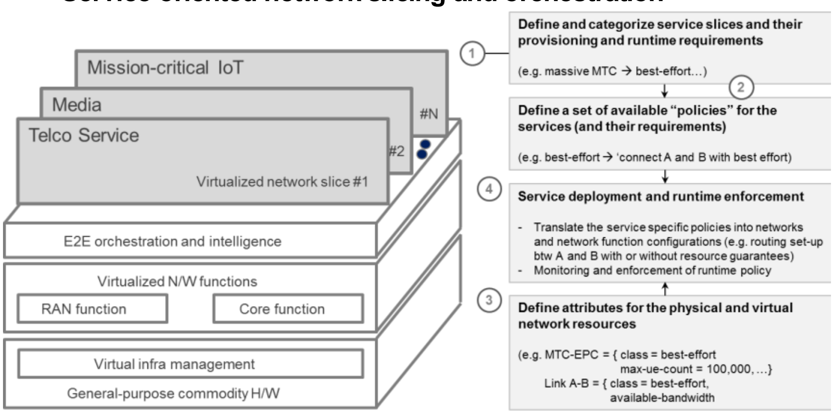


The three most distinguishing features of C-Core architecture are: 1) flat and distributed architecture based on user / control plane separated network functions, 2) service-oriented network slicing and orchestration, and 3) cost-efficient NFV decomposition and differentiation:

- **Flat and distributed architecture with separated user / control plane functions:**

Distributed edge cloud can also eliminate massive signaling and data traffic coming from the end devices and things by placing relevant control functions at the edge of the network close (possible at C-RAN locations) to the device. This allows edge cloud functions to process all traffics locally, for example, in B2B premises that need high level of security. On the other hand, the separated control plane functions will be aggregated / simplified / centralized for enhanced deployment, operations, and management.

- **Service-oriented network slicing and orchestration**



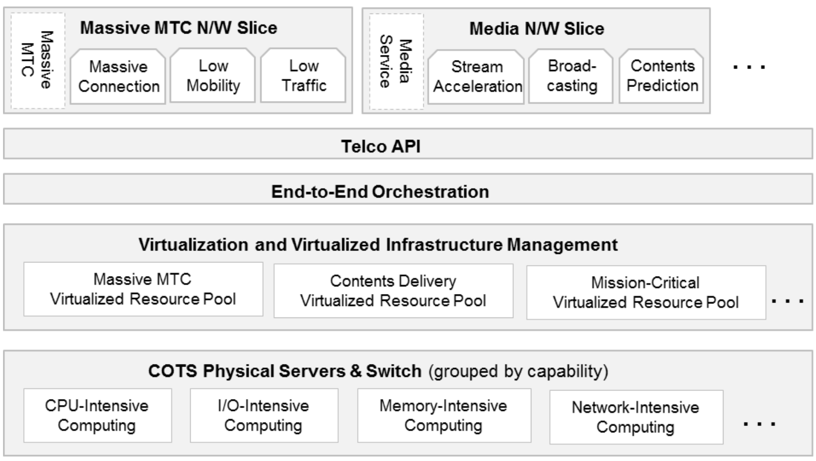
Distributed edge cloud can eliminate massive signalling and data traffic coming from the end devices and things *by placing relevant control functions at the edge of the network close (possible at C-RAN locations) to the device.*

In addition, ***the separated control plane functions will be aggregated / simplified / centralized for enhanced deployment, operations, and management.***

A network slice, namely "5G slice" is composed of a collection of 5G network functions and specific RAT settings that are combined together *for the specific use case or business model.*

NGMN 5G white paper [4] defines "5G slice" as follows: A network slice, namely "5G slice" is composed of a collection of 5G network functions and specific RAT settings that are combined together for the specific use case or business model. Thus, a 5G slice can span all domains of the network: software modules running on cloud nodes, specific configurations of the transport network supporting flexible location of functions, a dedicated radio configuration or even a specific RAT, as well as configuration of the 5G device. Not all slices contain the same functions, and some functions that today seem essential for a mobile network might even be missing in some of the slices. The intention of a 5G slice is to provide only the traffic treatment that is necessary for the use case, and avoid all other unnecessary functionality. The flexibility behind the slice concept is a key enabler to both expand existing businesses and create new businesses. Third-party entities can be given permission to control certain aspects of slicing via a suitable API, in order to provide tailored services.

Given various "5G slice" types, how to efficiently lifecycle-manage the slices becomes important. A policy-based end-to-end orchestrator manages both the virtualized network resource profiles and service requirements for each 5G service. Upon the request to create a 5G slice, the orchestrator then finds appropriate virtualized network resources for the slice from end-to-end perspective, given the requested service and the service requirements. The aforementioned C-Core architecture can be redrawn from the perspective of 5G slice and orchestration as shown below.



- **Cost-efficient NFV decomposition and differentiation**

As different 5G services have very diverse service requirements, it becomes clear that the virtualized resource may selectively be used to implement the corresponding 5G slices in a cost-efficient manner. For this, we expect that *there will be differentiated versions of a given function (e.g., cost-efficient version with minimal function set, premium version with costly features for high reliability and performance, etc.) for the operators and customers can use to build their slices*. Also, a given network function will also be decomposed to the smallest functional units (each with a price), and will dynamically be picked and chosen to make up a version that best fits the user's service requirements.

The following table summarizes the C-Core guidelines.

Aspect	***ID***	***D***	***L***	***C***	***Q***	***A***	***E***	***Guidelines***
IMP	CCG.1	○	○	○			○	Support of differentiated grouping and management of infrastructure based on hardware capability
	CCG.2			○		○	○	All network functions virtualization
	CCG.3	○	○	○				Virtualized C-Core functions shall satisfy 5G KPIs specified by ITU-R WP5D and potentially other industrial organizations such as NGMN
	CCG.4				○	○	○	Capability and policy aware VIM, VNF Manager, and Orchestrator
	CCG.5			○			○	Lean and decomposed NFV
	CCG.6				○	○	○	Carrier-grade per-DC orchestrator.
	CCG.7						○	Linear performance increase with the amount of overall resources allocated
	CCG.8				○	○	○	Support of (anti-)affinity of compute, network, storage, and location for HA, DR, and performance enhancements
	CCG.9				○	○	○	Support of seamless inter-DC migration and failover
	CCG.10				○	○	○	Support of inter-DC resource pooling
	CCG.11					○	○	Dynamic inter-DC bandwidth, connectivity, routing, and QoS management
DEP	CCG.12						○	Support of incremental upgrades
	CCG.13				○		○	Support of service capability exposure function – Telco API
O&M	CCG.14				○	○	○	Support of automatic, manual, and policy-based lifecycle management of a given network, network function, and

Requirements

D: Increased Data Rate
L: Reduced Latency
C: Massive Connectivity
Q: Guaranteed QoE
A: Higher Availability
E: Higher Efficiency

As different 5G services have very diverse service requirements, it becomes clear that ***the virtualized resource may selectively be used to implement the corresponding 5G slices in a cost-efficient manner****.*

To efficiently support the services and guarantee service QoE, ***there will be differentiated versions of a given function*** *for the operators and users can use to build their slices.*

								network service
	CCG.15				○	○	○	Means to achieve full physical and virtual infrastructure visibility on network, network functions, service and traffic dynamics for QoE management and efficicient troubleshooting

4.4.2. C-Core implementation guidedlines (CCGs)

4.4.2.1. Implementation aspect (IMP)

The physical infrastructure shall be based on All-IT COTS hardware.

*However, **there will be differentiated server groups with different capabilities and costs**.*

Eventually, the virtualized infrastructure must also be programmable (e.g., software-defined infra.)

[CCG.1] Support of differentiated grouping and management of infrastructure based on hardware capability

- The 5G core network functions shall run on all-IT infrastructure. The main goals are:
  - Overall TCO reduction
  - Ease of maintenance and management
  - Higher scalability and shorter TTM
- There shall be several groups of all-IT equipment managed separately, depending on their capability. For example, high-end premium COTS servers with costly features (e.g., CPU and NIC for data plane acceleration support, etc.) may be categorized as a “premium servers” group and used to provide premium services such as public safety network or HD voice service. The followings are examples of how the servers may be categorized:
  - “CPU intensive” group for control plane network functions
  - “memory intensive” group for control plane real-time DB functions
  - “storage intensive” group for control plane non real-time DB functions
  - “network intensive” group for data plane network functions

  The infrastructure shall be designed in an expandable manner to accommodate more server groups as needed in plug-and-play manner.

[CCG.2] All core network functions virtualized

- All network functions in C-Core shall be virtualized. The main goals are:
  - Cost efficiency (i.e., TCO reduction)
  - Flexibility and programmability, leading to operational efficiency
  - Time-to-market reduction
  - Energy efficiency

*For better programmability of the overall network, **all network functions in C-Core shall be virtualized.***

[CCG.3] Virtualized C-Core functions shall satisfy 5G KPIs specified by ITU-R WP5D and potentially other industrial organizations such as NGMN

[CCG.4] Capability and policy aware virtualized Infrastructure Manager (VIM), VNF Manager, and Orchestrator

- The virtualized infrastructure manager (VIM), VNF manager, and orchestrator shall keep a complete record on the characteristics, capabilities, possible configurations of all physical and virtualized resources, and the underlying physical and virtual networks
- The abovementioned records shall be managed using standardized format specified by ETSI NFV, TOSCA, YANG, and etc.

[CCG.5] Lean and decomposed NFV

- A given virtualized network function (VNF) shall be designed and implemented such that the network function can be offered with various options, where each option is associated with certain cost. For example, a vMME network function for public safety may come with high availability feature may be more costly than a vMME network function without the high availability feature. This approach will become quite desirable for massive IoT where the network functions may only be 'best effort' and where the overall cost reduction is more important.
- The control plane shall be separated from data plane
- The database (e.g., session database) shall be separated from other functional units as much as possible

*To support diverse services cost-effectively, **the existing network functions must be re-designed (or leaner) per different service types with differentiated capabilities** and **the underlying infrastructure must be built to support it accordingly**.*

The network functions shall be decomposed for enhanced quality of

- The communication amongst the decomposed functions of a given VNFs and VNFCs shall be standard-based (e.g., OpenFlow, other forms of SDN, etc.)

experience and cost-effectiveness.

[CCG.6] Carrier-grade per-DC orchestrator, where the per-DC orchestrator shall:

One of the most important entities in the programmable 5G system is the orchestrator. Different network domains should expose a set of APIs to the orchestrator.

- Offer End-to-end lifecycle management within the DC (on-board, instantiate, update, auto/manual scaling, heal, terminate, etc.) of various network services
- Offer management of network service catalog, service slice, and slice policy
- Offer northbound APIs to the users (e.g., operators and customers), such that the users can lifecycle manage the on-boarded services
- Offer centralized control of transport networks (i.e., transport SDN controller) for traffic management and control of different QoS level across different transport networks within and at the edge of DC
- Offer general orchestration functions as described in ETSI NFV MANO specification, at the "network service" level
  - NFV Orchestration: orchestrator shall provide functions as described in ETSI NFV MANO specification
  - Network OS: Network OS shall abstract necessary network resources and present them to the E2E and NFV orchestrators for automated network service lifecycle management (e.g., abstraction and offer intuitive network programmability). Network OS shall automatically detect network failures and automatically recover from the failure with minimal detection and recovery time (e.g., less than 1 second) if there are alternate available network paths for the services. Network OS shall be carrier-grade with built-in redundancy (with service availability greater than 99.9999%)

[CCG.7] Linear (as least) performance increase with the amount of overall resources allocated

The network functions should eventually be redesigned towards becoming virtualized

- Overall performance must increase linearly with the amount of allocated virtualized resources. Assume that a given function processes Z transactions/sec when running with X CPU and Y memory. When running with 2X CPU and 2Y memory, the function shall process at least 2Z transactions.

environment friendly and service-oriented*, as opposed to a simple porting of the legacy software.*

[CCG.8] Support of (anti-)affinity of compute, network, storage, and location for HA, DR, and performance enhancements

[CCG.9] Support of seamless inter-DC migration and failover

[CCG.10] Support of inter-DC resource pooling

[CCG.11] Dynamic inter-DC bandwidth, connectivity, routing, and QoS management

4.4.2.2. Deployment aspect (DEP)

[CCG.12] Support of incremental upgrades.

That is, the overall hardware or its components (e.g., CPU, memory, disk) shall be partially upgradable without being completely replaced.

[CCG.13] Support of service capability exposure function – Telco API

- APIs shall be offered to the users (e.g., operators, customers) for Network-as-a-Service and Infra-as-a-Service within the boundary of not compromising privacy and security of the network and data
- The APIs shall be highly secure and robust with built-in DDoS prevention mechanisms
- The APIs shall be offered through intuitive interfaces (e.g., web-based)

Apart from the internal APIs to the orchestrator for OA&M purposes, there will be a set of APIs defined from the orchestrator externally to enable 3rd party driven innovative value-add services.

4.4.2.3. Security aspect (SEC)

N/A

4.4.2.4. Operations & management aspect (O&M)

[CCG.14] Support of automatic, manual, and policy-based lifecycle management of a given network, network function, and network service.

The network, network functions, and services must provide sufficient flexibility and programmability for automated operations and management.

[CCG.15] Means to achieve full physical and virtual infrastructure visibility on network, network functions, sevice, and traffic dynamics for efficient QoE management and troubleshooting. As some of the traffic (e.g., east-west traffic of two VMs within a single physical host) may be hidden from the previous monitoring mechanisms, the overall infrastructure must provide a means to effectively monitor all dynamics and traffic even in the virtualized environment.

4.4.2.5. Additional functions aspect (ADD)

N/A

4.4.3. Other C-Core implementation considerations (CCCs)

- Open source software

  Cloud Core and its components will be designed to easily embrace open source software, as open source software tend to be one of cost-efficient approaches to implementing innovative services in timely manner. Examples of open source functions include: open source Telco functions (such as openEPC, etc.) as well as IT functions. The open source Telco functions (probably with best-effort quality) are expected to play greater role for massive IoT services, which may be implemented at a massive scale but at a 'best-effort' quality (and therefore, cost-efficiency overrules the strict QoS that must be provided).

- Open platform standards (for performance acceleration, etc.)

  Virtualized network function should adhere to open platform standards to a certain extend so that it can provide interoperability across various NFV solutions and infrastructures. An example of open platform standard for NFV is OPNFV (Open Platform for NFV). Such open platform may help reducing the overall testing time before deployment (e.g., multi-vendor testing time reduction) as well as accommodating different vendors.

- Carrier-grade SDN solutions (e.g., Open Network OS)

  SDN will play an increasingly important role in 5G for automated QoE management, working tightly together with NFV orchestrator. The SDN related functions (such as SDN controllers and switches) shall also be carrier-grade. Segment routing may be considered

Other implementation considerations for C-Core include:

- *Open source software*
- *Open platform standards*
- *Carrier-grade SDN solutions*
- *Incremental network evolution and interworking with physical / legacy equipment*
- *Network visibility for network diagnosis and troubleshooting*
- *Intelligence and network correlation & analytics*
- *Network programmability*
- *Light-weight OS-level*

as an option to guarantee QoE in the transport.

virtualization

- Incremental network evolution and interworking with legacy physical equipment

  Virtualized network function shall interwork seamlessly with legacy physical equipment for incremental network evolution support

- Network visibility for network diagnosis and troubleshooting

  Virtualized network functions (VNFs) tend to hide themselves within their containers (e.g., east-west traffic between VMs within the same host). Therefore VNFs and software for Cloud Core should provide a mechanism to make relevant monitoring, troubleshooting, and debugging information visible to the element management system (EMS) or network management system (NMS) for easy and intuitive network monitoring, diagnosis, management, and troubleshooting

As much information that were previously visible become invisible (such as east-to-west traffic), ***there needs an effective means to monitor the hidden traffic and performance in the new 5G virtualized environment.***

- Intelligence and network data correlation & analytics

  Virtualized network function or service should provide mechanism to deliver meaningful data (e.g., performance KPIs) to an analytics engine. The data shall be processed in (near) real-time for E2E performance management of the functions and services.

- Network programmability

  Virtualized function or service and network should provide mechanisms (e.g., APIs) to programmatically configure its parameters dynamically.

- Light-weight OS-level virtualization (e.g., Containers)

  Light-weight OS-level virtualization technologies should also be considered as an option for virtualization, as they may provide higher performance. In such cases, light-weight virtualization containers shall be standard-based and also the trade-offs (e.g., less security) should be studied thoroughly before being added to the network.

Document Management

Document History

Revision	Date	Brief Description of Change	Approval Authority	Editor / Company
0.1	5/22/2015	Initial draft		Jonghan Park, Changsoon Choi, Minsoo Na / SK telecom
0.11	5/29/2015	Overall reorganization; Reflected comments on E2E guidelines		Changsoon Choi / SK telecom
0.2	6/1/2015	Reflected comments on C-RAN guidelines		Minsoo Na / SK telecom
0.21	6/1/2015	Overall reorganization; addition of figures and tables		Sungho Jo / SK telecom
0.3	6/5/2015	Addition of C-RAN and C-Core considerations section		Minsoo Na, Jonghan Park, Changsoon Choi / SK telecom
0.4	6/5/2015	Overall polishing and typo corrections		Jonghan Park / SK telecom
0.5	6/8/2015	Reorganization of C-RAN guidelines into 5 aspects		Minsoo Na / SK telecom
0.6	6/15/2015	Overall refinement		Sungho Jo / SK telecom
0.7	6/16/2015	Reflected internal comments from 5G Tech Lab		Changsoon Choi / SK telecom
0.71	6/19/2015	Minor typo correction		Haesung Park / SK telecom
0.72	6/19/2015	Addition of acronym table		Yongseok Choi / SK telecom
0.73	6/20/2015	Minor typo corrections		Moonhong Kim / SK telecom
0.8	6/26/2015	Overall revision on C-Core, its architecture, and network slice description		Mingeun Yoon / SK telecom
1.0	6/30/2015	Overall refinement for v1.0		Jonghan Park / SK telecom
1.1	7/16/2015	Reflected internal comments from 5G Tech Lab; Redrew figures for improved consistency		Jonghan Park, Changsoon Choi, Minsoo Na / SK telecom
1.2	7/24/2015	Overall revision, especially on C-RAN		Minsoo Na / SK telecom
1.3	7/27/2015	Overall polishing		Changsoon Choi, Jonghan Park / SK telecom
1.31	8/1/2015	Added guidelines on inter-DC connectivity, routing, bandwidth and QoS		Jonghan Park / SK telecom
1.32	8/7/2015	Restructured the overall guidelines to be more MECE (mutually exclusive		Yongseok Choi / SK telecom

Revision	Date	Brief Description of Change	Approval Authority	Editor / Company
		and collectively exhaustive)		
1.33	8/11/2015	Added LPWA and new RAT TTI guidelines		Changsoon Choi, Minsoo Na / SK telecom
1.34	8/12/2015	Overall polishing		Sungho Jo, Jonghan Park / SK telecom
1.35	8/25/2015	Editorial revisions; reflected comments received internally from SK telecom Network Technology R&D Center		Sangsoo Jeong, Changsoon Choi, Minsoo Na, Jonghan Park / SK telecom